The Journal of Molecular Diagnostics, Vol. 26, No. 6, June 2024

ELSEVIER

the Journal of Molecular Diagnostics
jmdjournal.org

Microarray-Based DNA Methylation Profiling

Validation Considerations for Clinical Testing

Marco L. Leung,*† Zied Abdullaev,‡ Lucas Santana-Santos,§ John M. Skaugen,¶ Stephen Moore,‖ and Jianling Ji**††

From The Steve and Cindy Rasmussen Institute for Genomic Medicine, Nationwide Children's Hospital, Columbus, Ohio; the Departments of Pathology and Pediatrics,† The Ohio State University College of Medicine, Columbus, Ohio; the Laboratory of Pathology,‡ Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland; the Department of Pathology,§ Feinberg School of Medicine, Northwestern University, Chicago, Illinois; the Department of Pathology,¶ University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania; the Department of Molecular and Medical Genetics and Knight Diagnostic Laboratory,‖ Oregon Health & Science University, Portland, Oregon; the Department of Pathology and Laboratory Medicine,** Children's Hospital Los Angeles, Los Angeles, California; and the Department of Pathology,†† Keck School of Medicine, University of Southern California, Los Angeles, California*

Accepted for publication February 8, 2024.

Address correspondence to Marco L. Leung, Ph.D., The Steve and Cindy Rasmussen Institute for Genomic Medicine, Nationwide Children's Hospital, 700 Children's Dr., Columbus, OH 43215.
E-mail: marco.leung@nationwidechildrens.org.

Microarray-based methylation profiling has emerged as a valuable tool for refining diagnoses and revealing novel tumor subtypes, particularly in central nervous system tumors. Despite the increasing adoption of this technique in clinical genomic laboratories, no technical standards have been published in establishing minimum criteria for test validation. A working group with experience and expertise in DNA-based methylation profiling tests on central nervous system tumors collaborated to develop practical discussion points and focus on important considerations for validating this test in clinical laboratory settings. The experience in validating this methodology in a clinical setting is summarized. Specifically, the advantages and challenges associated with utilizing an in-house classifier compared with a third-party classifier are highlighted. Additionally, experiences in demonstrating the assay's sensitivity and specificity, establishing minimum sample criteria, and implementing quality control metrics are described. As methylation profiling for tumor classification expands to other tumor types and continues to evolve for various other applications, the critical considerations described here are expected to serve as a guidance for future efforts in establishing professional guidelines for this assay.
(J Mol Diagn 2024, 26: 447—455; https://doi.org/10.1016/j.jmoldx.2024.02.001)

Histopathologic features traditionally have been used to classify tumors according to the World Health Organization criteria. Over the past decade, many large-scale genomic studies have revealed various molecular features that are distinctive to particular tumor types. Since then, the World Health Organization has integrated molecular features to further classify and delineate tumor types and subtypes. A combination of molecular analysis and morphologic interpretation enhances the level of diagnostic precision and provides optimal prognostic and predictive information.

In recent years, microarray-based DNA methylation profiling has emerged as a valuable methodology for tumor classification; this technology is capable of determining DNA methylation status across the genome.[1] Previous studies have elucidated the clinical potential of DNA methylation—based classification for patients diagnosed with cancers, particularly central nervous system (CNS) tumors. In addition to recognizing known CNS tumor types, methylation profiling also has revealed novel subtypes and further refined existing known tumor entities. For example, methylation profiling could further subdivide non-WNT/non—sonic hedgehog (SHH) medulloblastoma groups 3 and 4 into eight distinct subgroups per the 2021 World Health Organization Classification of CNS Tumors.[1] In tumors that lack defining gene mutations or fusions, methylation profiling also can be applied. In different studies, methylation arrays resulted in a change in diagnosis in up to 18% of patients with CNS tumors.[2–5] As DNA-based methylation profiling becomes more integrated into clinical care for

Current address of S.M., LabCorp, Research Triangle Park, NC.

https://doi.org/10.1016/j.jmoldx.2024.02.001

patients with cancers, clinical laboratories are beginning to adopt this methodology into their laboratory practices. Because of its novelty, there have been no published guidelines describing the critical considerations for launching a clinical methylation profiling test. Furthermore, microarray-based methylation profiling is not described in the current College of American Pathologists checklists, and the College of American Pathologists does not provide any proficiency testing for this methodology. It can be challenging to find guidance materials for clinical laboratories seeking to validate and launch this test clinically.

This article compiles practical points for developing and validating methylation profiling in the clinical setting with the aim to improve patient care by sharing experiences using this methodology with the broader field of molecular pathology. The future of methylation profiling technologies and their applications to other diseases also are discussed.

Material and Methods

Six collaborators (molecular pathologists, laboratory geneticists, and bioinformaticians from different hospitals/institutions), each with experience and expertise in validating and reporting DNA-based methylation profiling tests on CNS tumors, collaborated for this study. This article offers practical discussion points and focuses on important considerations for validating this test in clinical laboratory settings.

Results

Herein, experiences in developing and validating microarray-based methylation profiling assay in the clinical laboratory setting are summarized (Figure 1). This article discusses the following aspects: technical validation considerations, tumor classifiers, automation of methylation workflow, quality control metrics, reportable findings (tumor classification, *MGMT* promoter methylation status, and copy number variants), and limitations of the assay.

Technical Validation Considerations

It is critical to determine the requirements for sample types, DNA input, and tumor cellularity for clinical testing. Although there are no general criteria for the required sample size in a validation study for methylation profiling, the typical sample size ranged from 40 to 100 among the laboratories of the authors in this study. Typical sample types used for methylation profiling include formalin-fixed, paraffin-embedded (FFPE) and fresh-frozen tissues. Although fresh-frozen tissues generally demonstrate better performance for many molecular testing applications, FFPE samples have shown robustness in methylation profiling.

The minimal DNA input required for methylation profiling ranges from 50 to 500 ng, with the most common input being 100 to 250 ng. Because samples often are archived in FFPE slides, only a limited amount of extracted DNA is available; validating a lower DNA input limit would

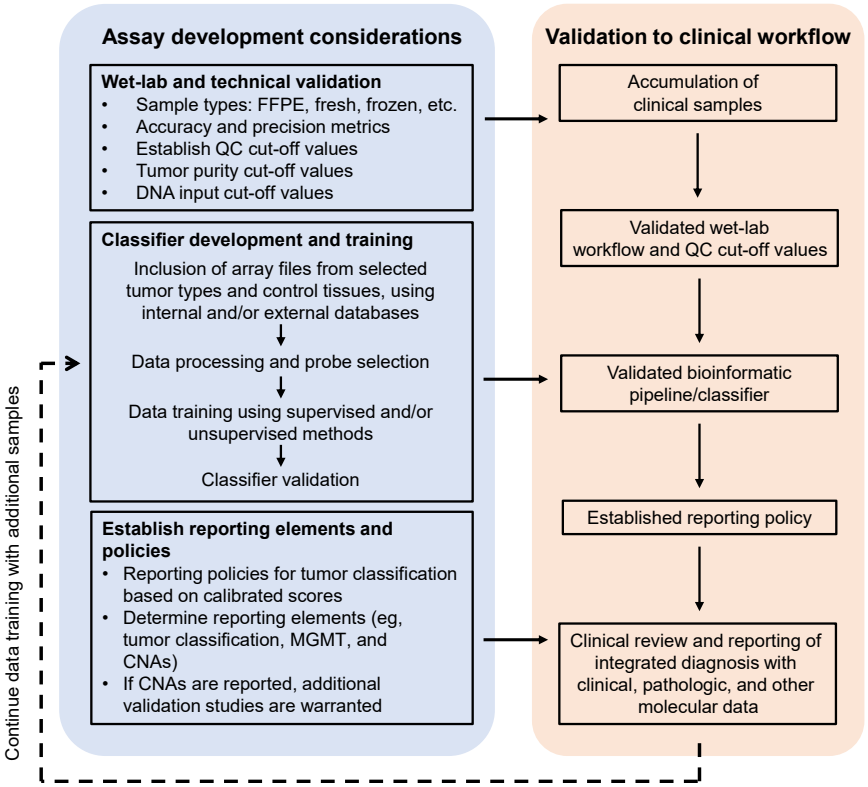

Figure 1 Assay development considerations and workflow. Various sample criteria and quality control (QC) metrics should be taken into considerations when validating the microarray-based methylation profiling assay and establishing the clinical workflow. Furthermore, the development and training of a tumor classifier require a sufficient number of array files from selected tumor types and control tissues to reach a high detection rate. CNA, copy number abnormality; FFPE, formalin-fixed, paraffin-embedded; MGMT, methylguanine methyltransferase.

alleviate this issue. Among the authors of this study, several have validated down to 50 ng.

Determining the tumor cellularity is important for methylation profiling, therefore it is critical to consult pathologists for their reviews on tumor purity before testing. Tissue with low tumor cellularity would expectedly have a weak tumor methylation signature, thus leading to lower diagnostic yield for the assay. Dilution studies play a crucial role in establishing the acceptance threshold for minimal tumor content percentage, which typically is set at 50% to 70%. Estimating tumor purity usually is performed through histology in routine clinical workflow but can present challenges, including requiring pathologist review and inaccuracies/subjectivities. In cases in which tumors have low tumor purity, laboratories may perform macrodissection techniques to obtain samples with higher tumor purity. Although a bioinformatics-based deconvolution method is available and has the potential to address this issue, it is not used commonly in routine clinical diagnostic settings.[6,7] It is important to note that the methylation classification tumor purity requirement may vary depending on the specific tumor type under investigation. For example, in the authors' experience, low-grade gliomas may require higher tumor purity than higher-grade tumors (such as medulloblastomas) to yield a classification with a high confidence score. Careful considerations of these factors is essential in the methylation profiling process to ensure accurate and reliable results for clinical applications.

The Illumina MethylationEPIC array (San Diego, CA) is a 4-day protocol with several 16- to 24-hour incubation steps (namely bisulfite conversion, DNA amplification, and chip hybridization), which may limit the start of the protocol (ie, the end of incubation may fall on an off day). Laboratories should consider validating longer incubation times (eg, up to 72 hours), which would provide greater flexibility for weekends and holidays.

Summary points to consider are as follows:

- Tumor purity is an important consideration when performing and interpreting DNA methylation profiling.
- Minimal tumor cellularity and DNA input should be established during validation.
- Validations of longer incubation times allows for greater flexibility and may improve turnaround time.

Tumor Classifier

Although externally developed classifiers, such as those developed by the German Cancer Research Center (DKFZ), may be available for clinical testing usage, some laboratories may elect to develop their own classifiers that may be more suitable for their patients' needs. Accessibility to a reference set of tumor samples is the most critical component for developing a tumor classifier. The performance of a classifier is heavily dependent on the inclusion of tumor subtypes used in the reference data set. The original DKFZ CNS tumor classifier had a wide variety of tumor samples ($n = 2801$) in their reference set, which yielded 91 methylation classes containing 82 tumor classes and 9 nontumorous classes in CNS v11b4 (*https://www.molecularneuropathology.org/mnp*, last accessed March 6, 2024).[2] The number and labels for each class/subclass in the reference sample set can affect the sensitivity and specificity of the assay. A higher number of samples used for each subtype likely would increase the sensitivity (ie, capability of classification) for the specific subtype. Conversely, if a specific tumor subtype is not represented, the classifier using the random forest algorithm would not classify that tumor subtype correctly, thus causing a decrease in sensitivity. The Capper et al[2] study described at least eight samples included in the subtypes of the DKFZ CNS tumor classifier (range, 8 to 143; median, 21). On the other hand, if a reference sample is labeled inaccurately, for example, if control tissue is labeled inaccurately as medulloblastoma, it would decrease the accuracy of the assay.

Understandably, it is challenging to include a high number of samples for rare tumor subtypes. Laboratories can either generate methylation array data from their internal tumor tissue bank or download publicly available data, such as the initial tumor array data ($n = 2801$) and validation data set ($n = 1104$) from Capper et al,[2] accessible on NCBI Gene Expression Omnibus (GSE109381 and GSE109379, *https://www.ncbi.nlm.nih.gov/geo*, last accessed January 31, 2024).[8]

The detailed development of classifier software is beyond the scope of this article. In brief, methylation array data can be analyzed utilizing two primary methods: unsupervised and supervised techniques. In unsupervised methods, data remain unlabeled throughout the analysis. To find the molecular structure of the data, computational techniques such as cluster analysis and dimensionality reduction, including methods such as t-distributed stochastic neighbor embedding and Uniform Manifold Approximation and Projection, commonly are used. These techniques facilitate the exploration of hidden patterns and relationships within the data, offering a deeper understanding of its underlying structure. This approach has the potential to resolve diagnostic uncertainties and provide an enhanced understanding of complex clinical entities. It is important to note that unsupervised methods do not provide the probability of sample classification (calibrated scores), and it should not be used alone in a clinical setting.

Conversely, supervised methods entail working with labeled data that represent predefined reference classes. In this approach, machine learning models are trained using the labeled data to predict the class of new unlabeled samples. The outcome of supervised analysis typically includes class scores or probabilities of class assignment for each sample. This information is crucial for classification and further in-depth analysis. Supervised classifiers, in particular, play a pivotal role in the clinical implementation of methylation profiling because they enable accurate sample classification

Table 1 Advantages and Challenges of In-House–Developed Classifier and Third-Party Classifier

Considerations	In-house developed classifier	Third-party classifier
Expertise requirement	Requires bioinformatic support to develop classifier software with a reference sample data set	Minimal bioinformatic support to install and integrate classifier software into the workflow
Software customization	Flexibility to adjust reference tumor sample set and improve classifications	Inability to modify classifier or refine classifications
Software location	Can be executed on a secure local server	May not be available for download and may require running samples on a third-party platform
Software ownership	Laboratory retains the ownership of classifier and does not need institutional–user agreements	May require institutional–user agreements and legal consults before use
Availability of reference sample data set	May be difficult to find adequate numbers and types of tumors to validate and generate a robust classifier	Robust classifier already generated

based on established criteria. Detailed descriptions of these methods can be found in previous literature.[9]

Advantages and challenges exist when using in-house–developed classifiers versus third-party classifiers (Table 1). Clinical laboratories should weigh their options based on their bioinformatics and computational expertise and capabilities, as well as their institutional regulations. For a laboratory with bioinformatic support to develop a classifier, which has access to reference data sets, and would like the flexibility to adjust and incorporate the reference data set to improve their classifier over time, an in-house–based classifier may be favored, particularly when institutional regulations may delay or pose challenges to signing a contract for a third-party license. Conversely, one may choose to adopt a robust third-party classifier for tumor classification when the laboratory lacks bioinformatic support or computational capability to train the classifier. For the latter, institutional–user agreements and legal consultations may be needed. Alternatively, some may opt to use third-party classifier software initially before developing an internal solution once they have gained experience with the assay.

It is important to note that classifiers used for methylation profiling are tied closely to a specific version of the microarray chip used for the data collection. In this context, in 2023, Illumina, the manufacturer of the widely used Infinium MethylationEPIC arrays, transitioned from version 1 to version 2, and the update brought substantial changes in its probe coverage. The notable enhancements were the removal of the poorly performing probes, and an increase in the number of CpG sites covered, expanding from previous version's coverage to more than 935,000 CpG sites in version 2. This update holds profound implications for classifiers that were developed using data from the earlier version 1 chip for the following reasons. First, probe coverage discrepancy: the shift from version 1 to version 2 of the chip includes a change in the set of CpG sites that are probed. This means that classifiers developed using version 1 data would not be compatible with version 2 data as a result of a significant discrepancy in probe coverage. Second, data compatibility: the classifiers rely on the specific features (CpG sites) provided by the microarray chip to make predictions. If these features change, as they did in the transition from version 1 to version 2, the classifiers would need to be retrained to work effectively with the updated chip data. Third, compatibility with existing analysis packages: existing software packages that were compatible with an earlier chip version might not work as a result of the changes in chip data annotation. Extensive code changes might be needed to fix these issues. Migration of version 1 to version 2 also may cause issues with existing libraries used for data normalization and copy number abnormality calling. Fourth, accuracy and generalization: classifiers trained on outdated chip versions may struggle to predict methylation patterns accurately when applied to samples generated using the newer chip version. This can lead to reduced accuracy and the potential for erroneous results in clinical or research applications. Fifth, research continuity: for clinical research studies that span chip version updates or when comparing data across different studies, it becomes imperative to ensure that classifiers are up to date. Otherwise, the data may not be comparable directly or may require complex normalization procedures.

In essence, the transition from one microarray chip version to another necessitates the re-evaluation and retraining of methylation classifiers to align with the updated data. Clinical laboratories must be vigilant about keeping their tools and models up to date, especially when working with rapidly evolving technologies such as DNA methylation arrays. This ensures the accuracy, reliability, and compatibility of methylation profiling results in the rapidly evolving field of epigenetics.

Summary points to consider are as follows:

- Both third-party and in-house–developed classifiers have advantages and challenges (Table 1); the decision may depend on the bioinformatic and computational

capabilities of the clinical laboratories, as well as any legal restrictions imposed by the institution.
- The sensitivity and specificity of the tumor classifiers are highly dependent on the number and variety of tumor samples, including their subtypes, in the reference data set.
- A significant change or update to the microarray chip would require re-evaluating and retraining the classifier to ensure the accuracy, reliability, and compatibility of the assay.

Automation of Methylation Workflow

As the demand for methylation analysis has increased, so has the need for efficient and reproducible sample processing methods. Automation plays a pivotal role in meeting this demand, offering advantages such as reduced hands-on time, improved accuracy, and increased throughput. The benefits of automation are as follows:

- Increased throughput: automation allows for the simultaneous processing of multiple samples, significantly increasing the number of samples that can be analyzed in a given time frame.
- Reduced variability: automation minimizes human error, leading to greater reproducibility in sample processing and more reliable results.
- Time efficiency: automated workflows reduce hands-on time, enabling clinical laboratories to focus on data analysis and interpretation rather than repetitive tasks.
- Improved data quality: by ensuring precise handling of samples and reagents, automation contributes to higher data quality and accuracy.
- Scalability: automated systems can adapt easily to changing workloads, making them suitable for both small-scale and large-scale studies.
- Traceability: barcode tracking and automated data logging enhance sample traceability, which is crucial for maintaining data integrity and regulatory compliance.

Challenges and considerations of the automation are as follows:

- Initial investment: the cost of acquiring and maintaining automation equipment can be substantial, necessitating careful budget planning.
- Workflow optimization: adapting existing protocols to automated systems may require workflow optimization and validation to ensure compatibility.
- Maintenance and training: regular maintenance and staff training are essential to keep automated systems running smoothly and efficiently.
- Sample integrity: automation systems must handle samples with care to prevent degradation or contamination.

The automation of methylation sample processing has the potential to enable researchers and clinicians to analyze epigenetic modifications with unprecedented efficiency and accuracy. As technology continues to advance, automated workflows will become increasingly attainable, making methylation profiling more accessible and robust for a wide range of applications. Proper planning, validation, and ongoing maintenance are essential to maximize the benefits of automation while minimizing potential challenges.

Summary point to consider: depending on the case volume, laboratories may consider adapting automation, such as liquid handling systems, to enhance throughput and scalability.

Quality Control Metrics

Because there is no quality control (QC) step during the sample processing protocol, it is crucial to assess sample quality during data analysis. Given the inherent characteristics of this assay, distinguishing whether a low confidence score in tumor classification results from suboptimal DNA/sample quality or the absence of a specific tumor type or subtype within the reference data set can be challenging.

To ensure the accuracy and reliability of methylation array profiles, it is essential to establish robust QC metrics as part of the validation study (Figure 1). These metrics serve as a critical checkpoint in the analysis pipeline. Various tools and resources are available for assessing data integrity before generating reports. Proprietary software such as Illumina's GenomeStudio and BeadArray Controls Reporters commonly are leveraged to scrutinize data quality comprehensively.

The GenomeStudio Methylation Module provides the number of CpG sites with reportable calls of a given array chip for each sample. The Illumina MethylationEPIC array v2 interrogates more than 935,000 CpG sites. As expected, if a sample fails to reach an adequate number of CpG sites with reportable calls there are insufficient data input for the classifier, which may lead to inaccurate tumor classification. For optimal downstream analysis, reaching at least 95% of detected CpG for fresh and frozen tissues, and 90% for FFPE samples, is recommended.

BeadArray Controls Reporter is an Illumina-developed software that can analyze intensity data (IDAT) files and assess the performance of control probes for BeadArray products. The methylation array chips contain multiple control probes that calculate various metrics such as FFPE restoration, staining, hybridization, and bisulfite conversion. The software has default thresholds for each control. For detailed thresholds for each control probe, refer to the Illumina BeadArray Controls Report Software Guide. Laboratories should review these control values across samples and sample types that have shown strong performance in the tumor classifier, and determine controls that should be considered as strict cut-offs or evaluate whether the thresholds need adjustment. When one or more QC metrics are outside the specified range, the data should be reviewed with caution. This evaluation of QC metrics

ensures the reliability of the methylation array data and enhances confidence in downstream analysis and interpretation.

Alternatively, R packages such as minfi (v1.34.0+; *https://bioc.cran.dev/packages/3.11/bioc/html/minfi.html*), meffil (v.1.3.6+; *https://rdrr.io/github/perishky/meffil*), and ewastools (v1.70+; *https://github.com/hhhh5/ewastools*) may be used for QC metrics assessment in lieu of GenomeStudio Methylation Module and BeadArray Controls Reporter for clinical laboratories with bioinformatics capabilities.[10–12] Regardless of QC software tools, QC metrics should be assessed during validation.

Summary point to consider: QC metrics should be established and incorporated into the assay protocol to ensure that a low confidence score for tumor classification is not the result of poor sample/DNA quality or processing issues.

Reportable Findings

Laboratories need to determine how to report tumor classifications based on methylation profiling (Figure 1). These classifications rely on calibrated scores, requiring the establishment of cut-off values to determine whether a tumor sample is considered a "match" to a particular type/subtype. When establishing a common calibrated score threshold, it is important to consider the trade-off between sensitivity and specificity. Capper et al[13] initially placed certain CNS tumors into eight methylation class families, each comprising several related classes. For cases within a methylation family, two scores typically are generated: one class score, which is not different from cases without a family, and one family score, representing the sum of combined class scores belonging to a methylation family. Introducing methylation families allows for a single threshold level across all methylation classes and families. The Capper et al[13] classifier uses a set threshold of 0.9. However, the threshold value should be assessed for every individual classifier; it should be calculated taking into consideration validation data and classifier performance. According to previous studies, the optimal trade-off between sensitivity and specificity (maximization of the Youden index) has been achieved with a calibrated score of 0.84.[13] Using the less-conservative cut-off value of 0.84 as a threshold may result in higher reportable findings in a clinical diagnostic setting. The subclass cut-off value of 0.5 has been widely adopted across laboratories, albeit chosen arbitrarily. For cases with calibrated scores between 0.5 and 0.84, there still may be some clinical value in providing suggestions from the classification, albeit with caution, and an integrated approach by incorporating clinical, histopathologic, and other molecular findings is recommended. Classifications with scores less than 0.5 generally are discarded.[13]

The methylation status for *MGMT* promoter also commonly is reported for patients with CNS tumors because *MGMT* promoter methylation has prognostic values for glioblastoma, as well as predictive values to response for radiotherapy.[14] If laboratories decide to report the *MGMT* methylation status as part of the overall assay, additional validation studies should be performed.[15] Typical readouts include methylated, unmethylated, or not determinable.

Summary point to consider: clinical laboratories should establish reporting policies for categorizing a patient's sample matching to a tumor type or subtype based on calibrated scores.

Copy Number Abnormalities

Copy number information can be extracted from the DNA methylation array data. Although not emphasized in this article, it is important to note that copy number abnormalities (can) could serve as diagnostic and prognostic markers for tumor classifications. For example, 1p/19q codeletion is a diagnostic marker for oligodendroglioma, C19MC amplification is a recurrent genetic abnormality for embryonal tumors with multilayered rosettes, and focal copy number alterations may be suggestive of fusion events, such as *KIAA1549*:*BRAF* fusion, which is seen commonly in pilocytic astrocytomas.[1] Examining copy number abnormality (CNA) plots generated from a methylation array also can serve as a QC metric, allowing for the assessment of data cleanliness. Similar to a conventional microarray for CNA detection, high-quality methylation arrays typically show a tightly clustered probe pattern, whereas low-quality runs may display a more scattered pattern. Additionally, the overall pattern of CNA, characterized by "up and down" alterations, can be used to establish correlations with other testing results, such as chromosomal microarray analysis. In cases in which a gene fusion is identified through alternative molecular methods, the methylation array potentially can facilitate breakpoint analysis if one or both genes involved in the fusion event are covered. Moreover, in cases in which methylation-calibrated scores are lower than optimal but still may provide clinical value to the overall diagnosis, CNA may be examined to determine whether they are consistent with the methylation classification. Validation of CNA detection for this platform is beyond the scope of this article. Analytical sensitivity and specificity of CNA detection should be evaluated if laboratories decide to report CNA as part of the assay.

However, limitations of CNA analysis using the methylation array platform do exist. One limitation is that the current classifiers, such as those developed by DKFZ, do not provide information on loss of heterozygosity, which is an important aspect of tumor genomics and typically can be obtained by routine array-based analyses. Additionally, the ploidy level may not be adjusted adequately, potentially

leading to inaccurate results, although relative losses and gains nevertheless can be assessed. Furthermore, CNA analysis may not distinguish between heterozygous and biallelic losses reliably, which can affect the interpretation of the findings and be affected by tumor purity.

Summary point to consider: CNA can provide additional information for tumor classification and may be reported, provided that validation studies are performed.

Assay Limitations

Although methylation profiling can help resolve diagnoses for tumors with challenging histologies and can place tumors into subclasses, it is not without limitations. The accuracy of tumor classification for this assay is dependent heavily on the composition of the training reference sample set used for the classifier. If a specific tumor type is not represented adequately in the reference sample set, a patient sample with that tumor type likely would yield no match, or a match to a different family/class with a low calibrated score. In other words, the classifier might fail to identify the tumor type correctly owing to the lack of sufficient training data for that specific category. Moreover, methylation profiling cannot provide specific genetic alterations definitively, including oncogenic driver events or druggable targets, even if the sample is matched to a tumor subtype with known specific mutations. Ancillary testing is needed to confirm these findings.

Diagnoses made based on methylation profiling alone are discouraged. Tumor classification by methylation profiling should be interpreted in conjunction with other testing modalities, including histology, cytogenetics, and molecular testing. There may be instances in which a patient's methylation results show discrepancies compared with other histology and molecular findings. Such discrepancies can arise because of various reasons, including the presence of novel tumor types that have not been characterized previously, or the utilization of different tumor sections for different testing modalities. Additionally, various scenarios, including the increased presence of tumor-associated macrophages, inflammation after treatment, undersampling, and specific tumor types such as low-grade glioneuronal tumors, may interfere with results from methylation profiling.[16] Such limitations should be spelled out in the clinical reports. It is important to note that microarray-based methylation profiling is an autonomous tool, not a diagnostician.[17] When the methylation profiling assay yields a classification with a high probability score that is discrepant from the initial clinical or pathologic diagnosis, a re-examination of the histopathology and additional ancillary assays may be warranted.[17]

Although it is not within the overall thrust of this article, it is worthwhile to mention that there currently is no specific Current Procedural Terminology code for methylation-based profiling. Nonetheless, alternative codes could offer a potential solution. Clinical laboratories planning to validate this test should take this into consideration.

Summary point to consider: the methylation profiling result should be interpreted with other testing, such as histology, cytogenetics, and other molecular testing.

Discussion

This work provides practical guidance for validating DNA-based methylation profiling for somatic diseases. With a methylation profiling assay that is validated thoroughly, this methodology can provide improved disease diagnoses and classification, reduce diagnostic inaccuracy, and improve patient outcomes. Until the College of American Pathologists publishes checklist items that describe the samples needed for validation or minimal assay performances for methylation profiling, the field will need to make its own best validation efforts to ensure accurate results and patient safety.

It is important to note that the considerations stated herein are agnostic to a specific classifier software. As software continues to improve, it is not practical to recommend the best classifier at a moment in time. Rather, recommendations are provided on validation in a broader context, for when a clinical laboratory decides to adopt this technology.

The classifiers have experienced continuous modifications with more tumor entities being recognized with each update in the past. Such an evolution will continue in the future. As the classifiers continue to advance, the diagnostic yield likely will improve as well. In a cohort of 269 patients, the DKFZ version 11b4 classifier resulted in classification with high probability scores in 66.4% of patients, whereas version 12.5 classifier, a more updated version, achieved high probability scores in 79.2%.[18] Knowledge of novel classes will expand with more reference samples containing rare tumor types, with the integration of other molecular results. Indeed, the study by Sturm et al[5] demonstrated that a multi-omic approach could increase diagnostic accuracy in a large proportion of patients. The expansion of the tumor entity poses potential challenges for individual laboratories to improve their classifiers continuously and to validate the updated versions in a clinical diagnostic setting. This challenge may increase, particularly when individual laboratories do not have access to the labeled reference samples to add to the training set. There may be a need to refine the machine-learning algorithm for data processing, emphasizing the importance of collaborative efforts in establishing a unified resource for the entire scientific community.

The application of DNA-based methylation profiling is poised for further evolution. Indeed, methylation-based classifiers have been developed beyond CNS tumor types, and early data show promising results in resolving difficult subtypes and improving outcomes for patients with sarcomas.[19] Although methylation profiling has been incorporated in clinical diagnosis by the World Health Organization only for CNS tumors, more well-powered studies likely will support the integration of this

methodology into routine patient care for other tumor types, such as soft-tissue tumors. Furthermore, DNA methylation array data also can be used to infer other types of data, such as chromatin accessibility, promoter methylation status (eg, *MLH1*), ethnicity, and immune cell content.[20–22]

With regard to methodologies, the Illumina Infinium MethylationEPIC array has been shown to be cost effective and high throughput for clinical methylation profiling, and, importantly, works extremely well with FFPE samples. Although current classifiers are based on array-based platforms, the declining cost of sequencing has sparked interest in exploring sequencing-based methylation, which may provide a more comprehensive assessment of the methylome.

Looking ahead, methylation profiling has the potential to extend beyond tumor profiling. Despite the challenges posed by small fragment size, low circulating tumor DNA concentration, and low circulating tumor DNA fraction, methylation profiling shows promise in liquid biopsy applications.[23,24] DNA methylation profiles from circulating tumor cells in plasma may reveal highly specific signatures for detecting and accurately distinguishing common primary intracranial tumors that share cell-of-origin lineages and can be challenging to differentiate using other methods.[24]

In conclusion, although challenges remain, methylation profiling holds the potential to improve patient care through precise and/or refined diagnoses and prognoses. The full clinical utility of this assay is expected to be achieved through collaborative efforts among clinicians, molecular pathologists/geneticists, and bioinformaticians. Together, these stakeholders can utilize the full potential of methylation profiling, driving advancements in personalized medicine and patient-centered care.

Author Contributions

All authors conceptualized the study; M.L.L., Z.A., and J.J. wrote the original manuscript; and all authors reviewed and edited the manuscript.

Disclosure Statement

All authors are affiliated with clinical laboratories that perform clinical microarray-based methylation profiling testing on a fee-for-service basis.

References

1. Louis DN, Perry A, Wesseling P, Brat DJ, Cree IA, Figarella-Branger D, Hawkins C, Ng HK, Pfister SM, Reifenberger G, Soffietti R, von Deimling A, Ellison DW: The 2021 WHO classification of tumors of the central nervous system: a summary. Neuro Oncol 2021, 23:1231–1251
2. Capper D, Jones DTW, Sill M, Hovestadt V, Schrimpf D, Sturm D, et al: DNA methylation-based classification of central nervous system tumours. Nature 2018, 555:469–474
3. Priesterbach-Ackley LP, Boldt HB, Petersen JK, Bervoets N, Scheie D, Ulhoi BP, Gardberg M, Brannstrom T, Torp SH, Aronica E, Kusters B, den Dunnen WFA, de Vos F, Wesseling P, de Leng WWJ, Kristensen BW: Brain tumour diagnostics using a DNA methylation-based classifier as a diagnostic support tool. Neuropathol Appl Neurobiol 2020, 46:478–492
4. Galbraith K, Vasudevaraja V, Serrano J, Shen G, Tran I, Abdallat N, et al: Clinical utility of whole-genome DNA methylation profiling as a primary molecular diagnostic assay for central nervous system tumors-a prospective study and guidelines for clinical testing. Neuro-oncol Adv 2023, 5:vdad076
5. Sturm D, Capper D, Andreiuolo F, Gessi M, Kolsche C, Reinhardt A, et al: Multiomic neuropathology improves diagnostic accuracy in pediatric neuro-oncology. Nat Med 2023, 29:917–926
6. Nissen E, Reiner A, Liu S, Wallace RB, Molinaro AM, Salas LA, Christensen BC, Wiencke JK, Koestler DC, Kelsey KT: Assessment of immune cell profiles among post-menopausal women in the Women's Health Initiative using DNA methylation-based methods. Clin Epigenetics 2023, 15:69
7. Zhu T, Teschendorff AE: Cell-type deconvolution of bulk DNA methylation data with EpiSCORE. Methods Mol Biol 2023, 2629:23–42
8. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A: NCBI GEO: archive for functional genomics data sets–update. Nucleic Acids Res 2013, 41:D991–D995
9. Yassi M, Chatterjee A, Parry M: Application of deep learning in cancer epigenetics through DNA methylation analysis. Brief Bioinform 2023, 24:bbad411
10. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA: Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 2014, 30:1363–1369
11. Heiss JA, Just AC: Identifying mislabeled and contaminated DNA methylation microarray data: an extended quality control toolset with examples from GEO. Clin Epigenetics 2018, 10:73
12. Min JL, Hemani G, Davey Smith G, Relton C, Suderman M: Meffil: efficient normalization and analysis of very large DNA methylation datasets. Bioinformatics 2018, 34:3983–3989
13. Capper D, Stichel D, Sahm F, Jones DTW, Schrimpf D, Sill M, Schmid S, Hovestadt V, Reuss DE, Koelsche C, Reinhardt A, Wefers AK, Huang K, Sievers P, Ebrahimi A, Scholer A, Teichmann D, Koch A, Hanggi D, Unterberg A, Platten M, Wick W, Witt O, Milde T, Korshunov A, Pfister SM, von Deimling A: Practical implementation of DNA methylation and copy-number-based CNS tumor diagnostics: the Heidelberg experience. Acta Neuropathol 2018, 136:181–210
14. Rivera AL, Pelloski CE, Gilbert MR, Colman H, De La Cruz C, Sulman EP, Bekele BN, Aldape KD: MGMT promoter methylation is predictive of response to radiotherapy and prognostic in the absence of adjuvant alkylating chemotherapy for glioblastoma. Neuro Oncol 2010, 12:116–121
15. Bady P, Sciuscio D, Diserens AC, Bloch J, van den Bent MJ, Marosi C, Dietrich PY, Weller M, Mariani L, Heppner FL, McDonald DR, Lacombe D, Stupp R, Delorenzi M, Hegi ME: MGMT methylation analysis of glioblastoma on the Infinium methylation BeadChip identifies two distinct CpG regions associated with gene silencing and outcome, yielding a prediction model for comparisons across datasets, tumor grades, and CIMP-status. Acta Neuropathol 2012, 124:547–560
16. Singh O, Pratt D, Aldape K: Immune cell deconvolution of bulk DNA methylation data reveals an association with methylation class, key somatic alterations, and cell state in glial/glioneuronal tumors. Acta Neuropathol Commun 2021, 9:148

17. von Deimling A: Methylation based classifications of human tumors. AMP 2023 Annual Meeting & Expo, November 16, 2023, Salt Lake City, UT. Rockville, MD: Association for Molecular Pathology, 2023
18. White CL, Kinross KM, Moore MK, Rasouli E, Strong R, Jones JM, Cain JE, Sturm D, Sahm F, Jones DTW, Pfister SM, Robertson T, D'Arcy C, Rodriguez ML, Dyke JM, Junckerstorff R, Bhuva DD, Davis MJ, Wood P, Hassall T, Ziegler DS, Kellie S, McCowage G, Alvaro F, Kirby M, Heath JA, Tsui K, Dodgshun A, Eisenstat DD, Khuong-Quang DA, Wall M, Algar EM, Gottardo NG, Hansford JR: Implementation of DNA methylation array profiling in pediatric central nervous system tumors: the AIM BRAIN project: an Australian and New Zealand children's haematology/oncology group study. J Mol Diagn 2023, 25:709–728
19. Koelsche C, Schrimpf D, Stichel D, Sill M, Sahm F, Reuss DE, et al: Sarcoma classification by DNA methylation profiling. Nat Commun 2021, 12:498
20. Benhamida JK, Hechtman JF, Nafa K, Villafania L, Sadowska J, Wang J, Wong D, Zehir A, Zhang L, Bale T, Arcila ME, Ladanyi M: Reliable clinical MLH1 promoter hypermethylation assessment using a high-throughput genome-wide methylation array platform. J Mol Diagn 2020, 22:368–375
21. Elliott HR, Burrows K, Min JL, Tillin T, Mason D, Wright J, Santorelli G, Davey Smith G, Lawlor DA, Hughes AD, Chaturvedi N, Relton CL: Characterisation of ethnic differences in DNA methylation between UK-resident South Asians and Europeans. Clin Epigenetics 2022, 14:130
22. Roy R, Ramamoorthy S, Shapiro BD, Kaileh M, Hernandez D, Sarantopoulou D, Arepalli S, Boller S, Singh A, Bektas A, Kim J, Moore AZ, Tanaka T, McKelvey J, Zukley L, Nguyen C, Wallace T, Dunn C, Wersto R, Wood W, Piao Y, Becker KG, Coletta C, De S, Sen JM, Battle A, Weng NP, Grosschedl R, Ferrucci L, Sen R: DNA methylation signatures reveal that distinct combinations of transcription factors specify human immune cell epigenetic identity. Immunity 2021, 54:2465–2480.e2465
23. Li W, Zhou XJ: Methylation extends the reach of liquid biopsy in cancer detection. Nat Rev Clin Oncol 2020, 17: 655–656
24. Nassiri F, Chakravarthy A, Feng S, Shen SY, Nejad R, Zuccato JA, Voisin MR, Patil V, Horbinski C, Aldape K, Zadeh G, De Carvalho DD: Detection and discrimination of intracranial tumors using plasma cell-free DNA methylomes. Nat Med 2020, 26: 1044–1047